The Project

David Tippett

Senior Developer Advocate

OpenSearch @ AWS

Personal Link Tree
dtaivpp.github.io/linkedme

Definition

OpenSearch is a community-driven, open source search and analytics suite derived from Apache 2.0 licensed Elasticsearch 7.10.2 & Kibana 7.10.2.

It consists of a search engine daemon, OpenSearch, a visualization and user interface, OpenSearch Dashboards, as well as a series functionality adding tools and plugins.

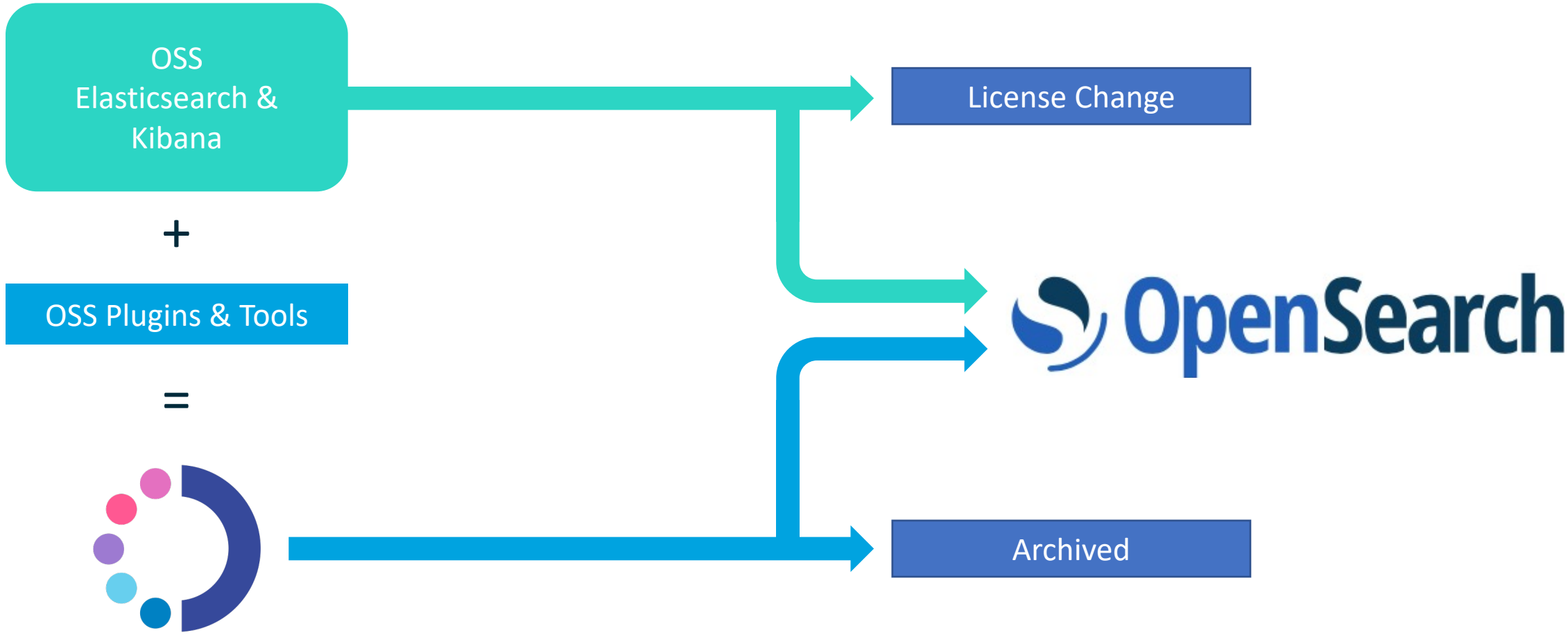
OpenSearch vs OpenSearch Project

	OpenSearch Project	OpenSearch
Specific	The GitHub repo	Just the search daemon
General	All components of OpenSearch	

What is OpenSearch intended for?

- Log, metric, and trace analytics
- Enterprise and general search

History


OSS
Elasticsearch &
Kibana
+
OSS Plugins & Tools
=
License Change
OpenSearch
Archived

License

Apache License, version 2 (ALv2)

- ✓ Use
- ✓ Modify
- ✓ Extend
- ✓ Embed
- ✓ Monetize
- ✓ Resell
- ✓ Make it part of your product and/or service

OpenSearch · OpenSearch Dashboards
All Plugins · All Tools
(Everything except the website: BSD 3 Clause)

Status

6/13/21		5/ 3/22	5/26/22		January 2023?
1.0 Release	**7 Minor releases**	**2.0 RC1**	**2.0**	**3.0 RC1**	**3.0**
✓	✓	✓	✓	⌛	

OpenSearch

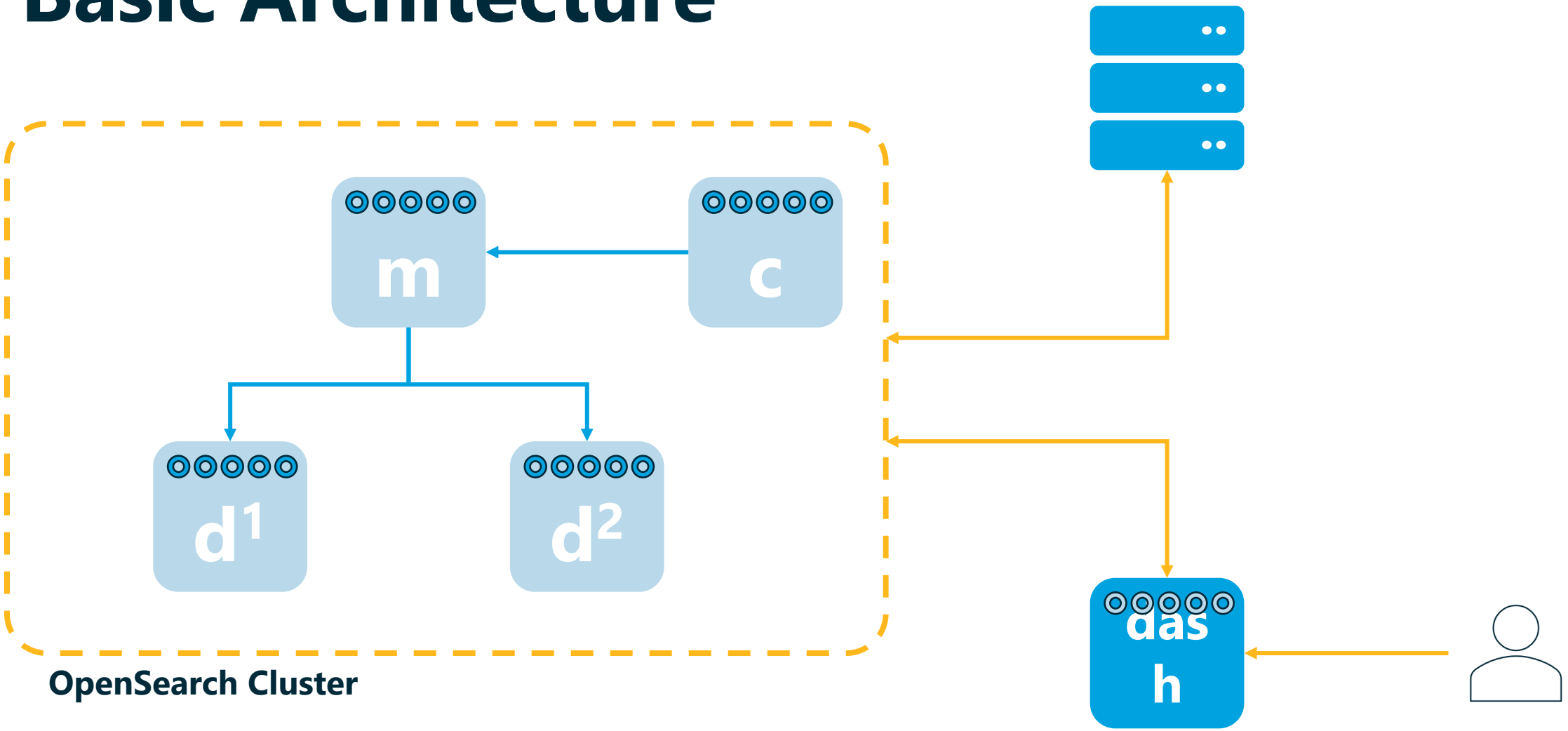
- Distributed search engine daemon
- Interact via REST API
- Performs indexing and storage of data
- Built in Java using Lucene
- Various types of nodes:

  Data · ~~Master~~/Main · Coordinating · Ingest · Machine Learning (v2.1)

OpenSearch Dashboards

- Browser-based UI and visualizations for OpenSearch
- Built charts and table representations
- Composes charts and tables into dashboards
- Interact with OpenSearch directly or use DQL
- Built in Typescript

Basic Architecture



Security and access control

- Node-to-node encryption
- Authentication
  - HTTP Basic, Active Directory, LDAP, Kerberos, SAML, OpenID Connect
- Role-based access control
- Index-, document-, & field-level access restriction
- Audit logging
- Dashboards multi-tenancy

Query languages

- Query DSL

```
{
    "_source" : ["name"],
    "query": { "range": { "RAM": { "gt": 16000 } } }
}
```

- OpenSearch SQL (including ODBC, JDBC)

```
SELECT name FROM computers WHERE RAM > 16000
```

- Piped Processing Language

```
source=computers | where RAM > 16000 | fields name
```

Alerting

1. Run a query on a schedule
    - *Run every 5 minutes (cron patterns)*
2. Trigger an alert when a condition is met
    - Example: *When n errors occur in a single hour*
3. Take an action on the alert
    - *Send a Slack message , fire a webhook, etc.*

Anomaly detection

Unsupervised machine learning (RCF) to find outliers in your data

- Create a detector
- Add features to the detector
- Observe results
- Set up alerts
- Adjust model as needed
- Manage detectors

Index state management

- Manage indices based on their properties
- Define *policies* to perform *actions* when the *state* in a *transition*
  - read only when > 30 days old
  - delete when 90 days old

Notebooks

- Create a narrative with visualizations and paragraphs
- Visualizations
  - Maps
  - Charts
  - Data
- Paragraphs
  - Markdown

Reports

- Sharing data and visualizations
  - PDF
  - PNG
  - CSV
- Can be created on-demand or on a schedule

Async Search

- Long running queries in the background
- Avoid any client disconnections
- Get back partial results

k-nearest Neighbors (KNN)

- Recommendations, image recognition, fraud detection
- Data is represented as vectors
- Hierarchical Navigable Small World graphs

Trace analytics

- Visualize trace data across complex architectures
- Supports OpenTelemetry
- Built to work with Data Prepper
  - Bring in data from various sources